DOCUMENT RESUME

ED 338 256 IR 053 787

AUTHOR Oddy, Elizabeth Carley
TITLE Kellogg Library and Archive Retrieval System (KLARS): Design and Development. Technical Report No. 6.
INSTITUTION Syracuse Univ., NY. Kellogg Project.
REPORT NO TR-1991-6
PUB DATE Aug 91
NOTE 44p.; For a related document, see IR 053 786.
PUB TYPE Guides - General (050) -- Historical Materials (060) -- Reports - Descriptive (141)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Access to Information; *Adult Education; *Archives; *Computer System Design; Database Design; *Full Text Databases; Higher Education; Hypermedia; Information Networks; *Information Retrieval; *Information Storage; Information Systems; Optical Disks; Optical Scanners; Records Management
IDENTIFIERS *Kellogg Project NY

ABSTRACT

This report provides a brief history of the design and development process that the Syracuse University Kellogg Project followed in creating the Kellogg Library and Archive Retrieval System (KLARS), an information storage and retrieval system designed to computerize archival research. This system was developed to preserve, organize, and facilitate access to Syracuse University's extensive English-language research collection of original resources on adult education, including manuscripts and archival documents, photographs, audiotapes, slides, and films. The report includes discussions of the issues faced during the design phase, a description of the system itself and its hardware and software, an outline of the project's solution to the problem of finding images of interest among the many images stored, and a description of the move to MacKLARS, a HyperCard adaptation of the KLARS system. The report concludes with a summary of the problems, contributions, and suggestions for those attempting to do something similar. (14 references) (Author/MAB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ED338256

Syracuse University Kellogg Project

Technical Report Series

KELLOGG LIBRARY AND ARCHIVE RETRIEVAL SYSTEM: DESIGN AND DEVELOPMENT

By Elizabeth Carley Oddy
Information Transfer Specialist
Syracuse University Kellogg Project
Syracuse, New York

Series Editors
Roger Hiemstra
Mary Beth Hinton

Technical Report No. 6
August 1991

IR053787

BEST COPY AVAILABLE

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Mary E. Hinton

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TR-1991-6

The Syracuse University Kellogg Project is designed to provide broad access to the University's outstanding collection of adult education materials using optical disk and computer technologies. The project, sponsored by the W. K. Kellogg Foundation of Battle Creek, Michigan, is based in the Adult Education Graduate Program. However, the Library and several other units of the University are involved. The Project's research and development efforts have resulted in new initiatives that we believe will benefit adult educators and others throughout the world.

This Technical Report Series is dedicated to disseminating knowledge regarding project initiatives, with the hope that sharing such information will improve the study and practice of adult education.

Copyright © 1991 Syracuse University Kellogg Project

CONTENTS

Abstract iv

Introduction 1

The Collections (Environment) 2

The Model 3

Design Considerations 3

Design and Development of the System 8

System Hardware and Software 14

Finding Images 17

Using KLARS 21

Move to MacKLARS 28

Suggestions 29

Contributions 33

Conclusion 34

References 35

ABSTRACT

This report gives a brief history of the design and development process the Syracuse University Kellogg Project followed in creating KLARS, which stands for Kellogg Library and Archive Retrieval System, an information storage and retrieval system designed to computerize archival research. The report includes discussion of issues faced during the design phase, a description of the system itself, and an outline of the project's solution to the problem of finding images of interest among the many images stored. It concludes with a summary of problems, contributions, and suggestions for those attempting to do something similar.

KELLOGG LIBRARY AND ARCHIVE RETRIEVAL SYSTEM: DESIGN AND DEVELOPMENT

By Elizabeth Carley Oddy

INTRODUCTION

Imagine that you are a researcher, preparing a history of the literacy movement in the United States. You have already combed the published literature for useful information. Now it is time to look at documents created by the various people and organizations who have actually done literacy work. You are aware that, in its George Arents Research Library for Special Collections, Syracuse University holds the papers of Laubach Literacy International and its founder, Frank Laubach, as well as the papers of Literacy Volunteers of America. These are unlike normal library materials in that they are unique collections of personal papers and organizational records.

To study these documents, you would have to travel to Syracuse and request permission to work in the special collections reading room, open only from 8:00 a.m. to 5:00 p.m. on weekdays. You would consult with the manuscripts librarian and look through finding aids to identify the collections you wanted to explore. Finding aids describe the person or organization that created the papers, and list groups of documents within the collection (a folder might contain "Correspondence, January-June 1962" from the Literacy Volunteers papers, for instance). You would not find the author/title/subject lists customary in libraries, nor would you find many things listed individually.

In archives, papers are stored in "archival boxes." A box holds either six or eighteen inches of stacked papers, depending on its type. You would request boxes of papers, then wait for someone to bring them out of the stacks. If the papers had been stored off the premises, you might have to wait a day or two to get access to them. The number of materials you could study at one time would be restricted, and nothing would be allowed out of the reading room. You might be able to bring a quiet portable computer into the room with you; otherwise you would take pencil notes. You could request copies of specific documents, but again you

would have to wait for them to be copied by a staff member, and pay a relatively high price.

This state of affairs has existed for a long time. Those who frequently use archive and manuscript collections accept the process and know how to make best use of time on study visits. But one man dreamed of changing the scenario. Professor Roger Hiemstra hoped in particular to make Syracuse University's valuable collections of adult education manuscripts more widely available and easier to use. The Kellogg Foundation funded his proposal, and in September 1986, the Syracuse University Kellogg Project began.

THE COLLECTIONS (ENVIRONMENT)

In addition to its extensive collection of relevant books and journals, Syracuse University is said to hold the largest English-language research collection of original resources--manuscripts and archival documents, photographs, audio tapes, slides, and films--in the field of adult education. The papers of approximately twenty professional organizations and of about thirty-five individuals constitute the bulk of the 900 linear feet of materials in the collection. Most of these materials are unpublished, many are unique. "Papers" is a comprehensive term that encompases correspondence, meeting minutes, lecture notes, snapshots, diaries, audiotapes of speeches, records of conferences, membership lists, by-laws--whatever the person or organization thought worthy of permanent storage. In addition to illustrating the development of adult education as a field, Syracuse's collection highlights educational aspects of many political and social phenomena of our time. Most of the material documents mid-twentieth-century North American activities.

Prior to the Kellogg Project, access to this unusual trove of research materials had been quite limited. One-third of the papers had not been taken out of the boxes in which they arrived in Syracuse. The thrust of the proposal to the Kellogg Foundation was to preserve, organize, and facilitate access to these resources.

THE MODEL

What if you could study unique manuscripts while sitting at a computer screen? We, the staff of the Kellogg Project, imagined a system in which you would be able to explore the subject area as well as the collections of papers themselves, using automated searching techniques. You then could use what you learned in that process to lead you to the groups of documents you sought. You would be able to call up facsimile images of documents--looking like photocopies but visible on your screen--to study. If you wanted a paper copy, you could send a command and a good laser copy would be printed for you. When you saw something of interest, you could take notes right away, in another window on your screen, without leaving the research work you were doing. And, best of all, you could do all of this at your own desk. We set about designing such a system.

In this technical report, I will begin with a brief history of the design and development process we followed in creating the Kellogg Library and Archive Retrieval System (KLARS), emphasizing the issues we faced during the design phase. Then I will describe the system itself and outline our solution to the problem of finding images of interest among the many images stored. To conclude, I will summarize problems we had, contributions we made, suggestions for those attempting to do something similar, and future possibilities.

DESIGN CONSIDERATIONS

Modifications to Current Practice

Providing access to archival documents in an image system turned out to be a very complex problem. We planned to incorporate what had been learned through many years of archival work, but felt we needed more specificity than the traditional emphasis on groups of documents could provide. Systems designed for retrieval of books and journals were specific enough, but weren't a good match with the kinds of documents we were describing. And imaging systems were designed for management of current office records, not for research. Somehow we had to bring these three approaches together (Keenan & Oddy, 1989). Before I describe our solution, let me explain some of the factors that affected our choices.

Comparison of Archives and Libraries

Although both libraries and archives collect and store documents for later use, there are important differences in how and why they do it (Rohfeld et al., 1990). Libraries generally collect published books and journals. Many copies of each book and journal issue are printed, and many libraries may have copies of the "same" ones. Each book or journal title is described individually in library records and catalogs, usually by author, title, and subject. The items can stand alone. Cataloging in one library does not differ in major ways from cataloging in another.

On the other hand, archives (or manuscript collections--there are technical differences between these terms, but I am using them interchangeably) usually hold unpublished, often unique, written or printed material. A particular collection is rarely available in more than one place, and then only by copying each document (e.g., using microfilm). Items are grouped. For example, committee correspondence from a certain period might be kept together, filed in an association's records under the name of the committee, and organized chronologically in batches. This grouping of information is done not only because of the sheer bulk of collections, but also because historians want to consult documents in context (Dowler, 1988).

Books and journal articles are written "about" something, to an unseen audience. Online catalogs, indexes, and abstracts connect the audience to available materials, along standardized lines (author, title, subject). Context is presented in documents by authors, along with the ideas themselves.

Archival documents, on the other hand, are paper leftovers of business accomplished, life lived. They are reflective "of" activity rather than "about" it, so they resist being categorized into subjects. It is often difficult to establish an author, and most documents have no titles. The original audience to which a document was written may have had shared experience, which could be alluded to without explanation. In short, researchers are looking in on something that was not intended for their eyes. This is the raw material of historical research, not the synthesized result. To understand what they are seeing, researchers must explore the context in which the documents were produced.

Knowledge of context is also required when searching for potentially interesting material. Which person or committee would be likely to have dealt with the subject you want? When was it being discussed? The researcher and the archivist are normally the sources of this contextual information. Researchers are generally held responsible for knowing the subject domain before attempting to work with original materials. Archivists and manuscript librarians have customarily made their considerable understanding of their own collections available to researchers, in reference interviews as well as in carefully prepared finding aids.

Finding aids list the physical locations of folders and boxes of records, but, more important, they include information describing the history of the papers and the people who created them, the kind of material that will be found in the collection, how the papers are arranged, etc. Researchers and archivists infer the likely location of useful information. Knowing about the organization or person who created the records provides clues to activities, and documents that result from those activities can be pursued from the clues. The focus is not on the documents but on the activity (Bearman & Lytle 1985-86; Lytle 1980a, 1980b; Pugh, 1982).

Thus, intellectual access to archive and manuscript collections has generally required a great deal of advance preparation by researchers, along with ready access to information stored in the heads of archivists. Finding aids that describe collections box by box, folder by folder, show users how documents are arranged. To get beyond this level to specific documents relevant to a researcher's interest requires understanding the environment in which the papers were created, a lot of browsing, and reliable hunches. We proposed to bring computer power to bear on this partly intuitive process.

Characteristics of Images

We assumed the presence of document images in our system. This meant that the searching task would require the system to locate not just a likely area for exploration, but to produce facsimiles of the documents themselves on the screen. To understand our decisions, it is important to know how images and machine readable text differ.

To create an image record, a page is scanned, or digitized, into the computer in a process very like photocopying, except that the product is an image on a screen rather than a paper copy. Imagine a very fine grid placed over the page, whether it contains text, a drawing, a photograph--the content doesn't matter. At each square in the grid, the computer analyses whether ink is or is not present. Image scanners vary in complexity, but in the simplest case this is an either/or decision. If there is enough ink in the square, the computer records a full cell or *bit;* if not, an empty cell is recorded. The result is a bitmap of the whole page. The important thing to note is that an image of a page of text is only a pattern of full and empty cells to the computer. The machine cannot "read" the characters, so no processing of the text is possible.

Most people are more accustomed to working with machine readable text . In this case, each character in a text is represented by a standard code (usually patterns of eight ones and zeros), and the codes, grouped into words, can be sorted, edited, searched, etc.

A problem we faced when designing our system was that, although images of the documents would be available online, the text in those bitmapped images would not be readable by the computer. If we wanted to be able to search and manipulate the textual content of documents, we would have to use an optical character reader (OCR) to translate the bitmap into character data. The OCR machine matches patterns found in bitmaps to its knowledge of shapes of characters and outputs its guesses in machine readable form.

Such translations could provide access to the actual content, as opposed to descriptions, of the digitized document images. Unfortunately, many of the documents in the adult education collections do not lend themselves to OCR, which works best on clearly printed material. Handwriting, fuzzy carbon copies, poor mimeographs, thermofaxes--many common mid-century methods of recording information--are of poor enough print quality to preclude optical character recognition. Even the most cleanly printed documents rarely translate perfectly. It would not be possible, therefore, to depend on OCR to provide access to the collections' contents, although OCR could probably provide a great deal of help in some situations.

Images take a great deal of space. In modern systems, complicated mathematical rules are used to compress the images and reduce required storage space, but it is always substantially more than the space required to store machine readable text. Imagine that you have a typed page of text, done in a font that gets ten characters to the inch. One letter *a,* taking up a square one-tenth of an inch per side, would create thirty times thirty bits of information, or 900 bits. It takes only eight bits to record the machine readable code for a single character. Compression ratios vary, depending on the content of the image, but even with substantial compression, images are storage grabbers. Our system would have to incorporate mass storage devices of some sort.

Structures for Storing Data in the Computer

In addition to combining images and OCR, we had to figure out how to organize our data for effective use. Because manuscript material traditionally has been organized very differently from books and journals in libraries, we could not just use a structure designed for library use. Authors, titles, and subjects are not good labels for retrieval of archive/manuscript material, as explained above. On the other hand, we could not simply employ the automation efforts already under way in the archival community, because those efforts concentrated on describing whole collections of papers. We would need to be much more specific when providing pointers to images.

To complicate things further, we felt that the rather rigid database management software generally available in image systems in early 1987 was inappropriate to our needs. Such software is well suited to storage and retrieval of information in standard formats, for which queries are relatively predictable. Many offices provide such environments. For instance, imagine an insurance application in which insurance policies, all following the same physical format, are captured as images and stored; standard information, such as policy number and policy holder's name, is made available for searching, and the kinds of queries put to the system are limited. Image system vendors paled when they saw the information stored in the adult education collections--assorted formats, widely varied quality, completely unpredictable and open possibilities for research approaches. For our system, we decided, text retrieval software would be an important addition to the database. Text retrieval, usually the choice in libraries, allows users to search for words in text

wherever they may be, and it can have a much less rigid structure, facilitating research. Unfortunately, no vendor was then marketing an image system with text retrieval capability adequate to our needs.

Our prototype, then, would need to combine these major components:

- Capture and storage of images of documents in the adult education collections,
- Translation of image bitmaps into machine readable code through optical character recognition,
- Retrieval of all machine readable text (translated image content and manually or electronically entered descriptive data).

DESIGN AND DEVELOPMENT OF THE SYSTEM

Choosing a Vendor

During the first few months of the Kellogg Project's life, Project Director Roger Hiemstra gathered a committee to develop a request for proposals (RFP) for distribution among potential vendors of optical disk systems. The project staff was not yet complete, so most committee members came from the wider campus--Anne Shelly and Ernest Sibert from the School of Computer and Information Science; George Abbott from the Syracuse University Library; Dana Cartwright from Academic Computing Services; and Dan Vertrees, the project's assistant director for computer and data integration. The committee came up with a very specific set of requirements for the future system. The RFP was not framed as a design, but rather as minimum capabilities any vendor's system would have to provide.

The group distributed the RFP to twenty-four optical system vendors in January 1987. Seven vendors responded with proposals; five of them came to Syracuse to present their suggestions to the committee, now enlarged by the addition of new project staff members Beth Oddy (information transfer specialist) and Terry Keenan (manuscripts librarian). The final three candidates came back for additional meetings during the last stages of decision making.

After a great deal of discussion, reference checking, and study, the committee selected Plexus Computers, Inc. of San Jose, California in April 1987. Many factors entered into the decision, of course, but the most compelling reason for choosing Plexus was their willingness to look for new solutions to the challenge we presented to them. Standard systems would be inadequate to the task. Plexus was the only company that proposed customizing software to meet our specific needs.

Contract

It took four months to hammer out a contract between Syracuse University and Plexus Computers, Inc. The university had recently been burned in a computer deal, and was extremely cautious and careful in negotiations. (Ironically, all that care was of no use when the worst--bankruptcy--happened.) Nonetheless, they did their best to protect the university's interests. On August 28, 1987 we signed the contract for the creation of Kellogg Library and Archive Retrieval System (KLARS).

Custom Software Specification

At the time the contract was signed, detailed specification of the system had not been worked out. None of us really knew what we were getting into. Project staff knew generally what they wanted users to be able to accomplish, and knew a good deal about the kind of information that would be kept in the system and how that information might best be organized. The Plexus team, initially consisting of Bob Marsh, chairman of the company, and Bob Nicholson, a software engineer, were experts on the hardware and more-usual software. It required a substantive joint effort to develop KLARS...

Plexus proposed, and the project accepted, the idea of dividing the design work into three phases. Phase one was intended to give us experience, rather quickly, with a prototype system. Much of the equipment would come within phase one. Data capture software would be standard, although data structures and retrieval software would be crude. Phase one would give us a chance to experiment with the equipment and to feed ideas into the design of phase two, our customized software.

Phase two would incorporate the full implementation of our image/OCR/full-text retrieval model, as well as two new pieces of equipment, the optical character reader and a jukebox for holding and manipulating fifty-one optical disk platters. Phase two software was expected to make it possible for staff members to accomplish the following tasks:

- Create digitized facsimile images of paper documents;
- Enhance those digitized images (bitmaps) to make them clearer and easier to read;
- Store them on optical digital Write Once Read Many (WORM) disks, which could each hold a gigabyte of data (a little more than 1,000,000,000 bytes);
- Create and enter data into a structure for organizing information about the documents;
- Convert image bitmaps into machine readable text using an optical character reader; and
- Index all substantive machine-readable words associated with images to make them available for word searching.

In phase two, researchers would be able to:

- View the images of documents on computer screens,
- Compare images on screen,
- View descriptions of groups of documents,
- Search for occurrences of specific words or combinations of words in those descriptions,
- Order facsimile prints of images,

- Access campus mainframe computers,
- Take personal notes on removable floppy disks, and
- Recall specific images from addresses, or markers, kept in personal notes.

Phase three was to be more experimental, testing feasibility of including digitized audio and color images in the system, and of providing access to index information at a distance over standard computer networks.

Phase One

Since the primary purpose of phase one was to become familiar with the hardware, the specification process was relatively simple. Phase one used a modified version of demonstration software already available at Plexus. Minimal description could be entered for images, so complex record structures were unnecessary. We anticipated using phase one for approximately five months, while phase two was defined and implemented.

Definition of phase one, a process begun while the contract was still under negotiation, was approved in September 1987. Most of the interaction between teams was accomplished by telephone or by mail. Discussion of system specifications also took place during contract-related visits to Syracuse by the Plexus team. The completed software and hardware were installed in Syracuse during the following December.

Staffing the system and housing it in existing space required a major effort, more demanding and time consuming than anticipated. A room in the library was rebuilt to system specifications, requiring the displacement of several library employees and construction of new space within an already crowded building. In addition to physical space, the effort required administrative time: staffing needs had to be defined and new positions described and filled, following guidelines of the library's union; lines of authority had to be worked into an existing hierarchical structure. Delays in accomplishing these simple sounding but rather difficult tasks caused a delay in use. Phase one became fully operational by March 1988. Because minimal customizing was involved, there were few bugs in the system.

Phase Two

Definition and design of phase two were entirely different and much more complex matters. The software would be unlike any Plexus had built before. Many people were involved. In addition to the people on the vendor selection committee, the Syracuse team consulted with others on campus including Jim Tompkins from Computer Network Services, and Don Marchand and Bob Oddy from the School of Information Studies. When they joined the project, Rae Rohfeld and Barbara Florini (adult education faculty) were added to the group. I was the Syracuse team leader. There were usually five people working on the Syracuse project in San Jose: Bob Marsh (later replaced as team leader by Gregg Stratton), Bob Nicholson (primary architect of the system who left after the design was approved), and software engineers Ken Ray, Don Julien, and Dennis McMahon. Although occasional visits would be imperative, an effective way of communicating frequently, easily, and at reasonable cost had to be established.

All of the Syracuse team members had access to computers connected to BITNET, an international telecommunications network linking universities. The Plexus engineers set up accounts in the SUN.COM network, similar to BITNET, but available to commercial enterprises. We were able to set up a group account here at Syracuse that allowed all of the team members, from both sides of the country, to read and respond to all interactions. This computer link largely eliminated problems of time and distance. Each team member could react, make suggestions, and comment at whatever time was convenient. Usually communication between the BITNET group and the SUN.COM group was instantaneous. Electronic mail became heavily used.

The design process began with conversations in Syracuse between Kellogg team members and Plexus staff. The Plexus group then proposed a specification for the system. Proposals were sent to all members of both design groups, and discussion took place using the joint account. Several drafts were similarly evaluated. The Plexus team visited Syracuse several times in the active design period for more thorough discussion. After the external system specification (written from the user's point of view) was complete, the Plexus team created an internal specification (from the programmer's point of view). From that document, software could be written.

Using this process, we were able to agree upon internal and external specifications for KLARS by February 1988. At that point, implementation of the design became the focus of attention, a job carried out by the Plexus team. The vendor anticipated a development time of approximately five months, but the task turned out to be considerably more complex than anticipated. Twelve months later (February 1989) a not-quite-finished version was installed in Syracuse. We agreed that it made sense to do the final debugging on our equipment rather than on slightly different equipment in California.

Final sign-off was expected by the end of March 1989. Unfortunately, in mid-March, Plexus Computers, Inc. filed for protection under Chapter 11 of the federal bankruptcy code. We had an unusable, incomplete system on our hands. Phase one, which had worked, had been dismantled when phase two was installed. After a summer of negotiations, a completion contract with Harshman Consulting, Inc. (HCI) of Santa Clara, California was signed on August 4, 1989. The new contract eliminated the originally proposed phase three and most of the planned custom software support, but otherwise was very similar to the original contract with Plexus. Dennis McMahon was retained from the Plexus programming team, but was only available to work on KLARS part-time.

Progress on completion was very slow. As the project's time allotment slipped by, plans for system use were gradually reduced. Many, many bugs and hardware problems plagued KLARS and a solution to one problem would only uncover a new problem. A fully indexed, working system eluded us until April 1991, only four months before the end of Kellogg Project funding. University support for the project was insufficient to allow KLARS' survival after the end of the grant, so the prototype was not developed further.

Training

The contract with Plexus Computers specified that Syracuse would be responsible for developing training materials for the system. Plexus was responsible for delivering complete internal and external specifications, as well as test scripts (which outline the steps required to test various options). A committee of Kellogg staff members, led by Jane Hugo, drafted training documents both for staff and for researchers. Composition of the committee varied over time, depending on which

aspect was being dealt with, but it included Beth Oddy, Terry Keenan, Kellogg/Library manuscript processors Molly Fulton and Peter Hammon, system administrators Dianne Hagan and later Yuming Tung, special collections curator Mark Weimer, and Rae Rohfeld and Roger Hiemstra, who offered their perspectives as adult education scholars.

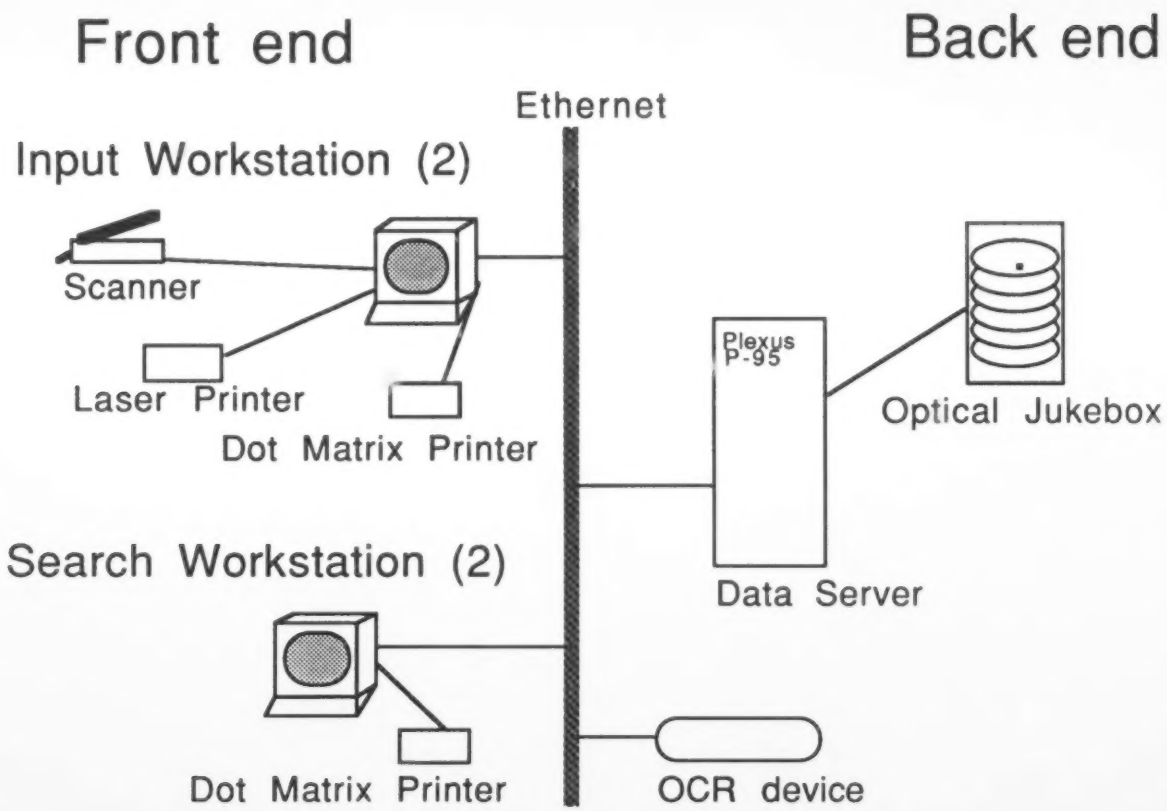
SYSTEM HARDWARE AND SOFTWARE

Hardware chosen for KLARS included a central processing unit (the main computer) connected by a local ethernet electronic network to four user workstations, two configured for staff work, and two for the use of visiting researchers. Also on the network were two high-speed, high-resolution laser printers and two digitizing scanners (for the use of staff), four small dot-matrix printers, a jukebox in which laser WORM disks were stored, and an optical character recognition device. These various pieces of hardware were grouped into *back end* and *front end* components (see Figure 1).

The back end was based on UNIX operating system software, and it used extended versions of Informix and SQL for managing and manipulating data. The front end workstations were essentially enhanced IBM PC XT clones, using MS-DOS and Microsoft Windows software. Plexus' own image processing software was incorporated as well. Special *socket library* programs allowed the front and back ends to pass instructions and data back and forth.

The system was designed to keep the back and front ends independent of each other; in other words, it would be possible to substitute new equipment (and the software to implement it) on one side without disrupting the existing equipment and software on the other side.

Figure 1. Diagram of Hardware configuration for the KLARS system.



In designing a system for the Kellogg Project, Plexus engineers built custom modifications onto the basic arrangement described above. For the back end, the major addition was information retrieval software. After much discussion, research, and negotiation, Plexus leased the PLS retrieval system from Personal Library Software, Inc. for use in our system. PLS would take care of the search and retrieval of information; Plexus would build their own interface onto the combined PLS and Plexus system. (See Figure 2 for a list of the hardware components and the software that ran on them.)

Figure 2. Hardware components and related software.

1 Plexus P/95 DataServer (central processing unit)
- 16 megabytes of main memory
- Operating system: Unix 5.2 (with BSD 4.2 enhancements)
- Used XDP DataManager (Plexus' version of Turbo Informix) Extended to allow storage of very large data objects)
- Supported by 572 megabytes of magnetic disk storage

1 Cygnetics Jukebox, model 3070A
- Included two 12-inch Optimem optical disk drives
- Used 12-inch Optimem WORM disks, 1 gigabyte per side
- Stored up to 51 optical platters

4 Plexus XDP Workstations (2 for input, 2 for searching)
- Wyse 286 IBM clones with 19" Sigma Designs high-resolution monitors
- All ran MS-DOS and Microsoft Windows software
- All contained Plexus image co-processor boards
- Communicated with university mainframe computers
- Each included one 5 1/4 inch floppy disk drive
- Each was attached to a small Toshiba dot-matrix printer, model P321SL

2 Fujitsu digitizing scanners, model 3094A
- Included automatic document feeders (up to 50 pages)
- Could scan legal, letter, A4, A5, B4, B5 sized paper
- 200, 240, 300 or 400 dot-per-inch (dpi) resolution

2 Fujitsu Laser printers, model M3727
- Printed 12 pages per minute at 300 dpi resolution

1 Calera 9000 OCR Server
- Omnifont character reader
- Read typewritten, typeset, laser-printed, dot-matrix styles
- Coped with proportional or uniform character spacing

All of the above were connected by a local Ethernet network.

Interface

Predictably, since the underlying workstation software was Microsoft Windows, the interface used multiple windows, pull-down menus, and a mouse. Specifying an MS-DOS operating system meant that users would work with directories and lists. Online help was not available.

System users would be assigned UserIDs and passwords. Passwords would be, of course, under the control of the user. The system manager would set access codes for each user. Only staff, for instance, could write to the optical disk. Appropriate training would be required of all users. KLARS was not designed for walk-up use in public space.

The user interface for searching was provided by Plexus, although the actual search software was from PLS. The system did a standard search on terms input by a user. The usual enhancements were available (proximity searching, field restriction, truncation, wild cards, range operators for numeric values). In addition, searchers could use previous queries or texts of previous retrievals in new queries. After the preliminary retrieval was complete, PLS ranked the results according to frequency of occurrence of search words. Users, then, could hope to find the most relevant material at the top of the list of retrievals.

FINDING IMAGES

Image bitmaps only record the presence or absence of ink. To enable users to locate specific images, we had to describe the documents in some way and let users look through the machine readable descriptions. Since we were working with archival material, it was not reasonable to try to identify each document individually (as one might do with books). We were faced with the difficult problem of deciding what information to collect and how to label and describe it in order to give users the best chance of finding what they sought.

It was obvious early that the data structure would have to include information from at least two levels. Small groups of documents labeled specifically enough to lead to individual pages would be needed, but would be far more useful if their relationships to the collections from which they came were made explicit. We

would therefore need descriptions of small groups of documents (set records) and descriptions of whole collections (collection records).

However, a consulting visit from David Bearman, editor of the quarterly journal *Archives and Museum Informatics* and a well-known expert on automation in archives, made it very clear that description only of these two kinds of data would severely limit the system's usefulness. In response to ideas discussed in that visit, we designed a data structure that also included information about careers of individuals important to the field of adult education (person records), information about important organizations and professional associations (organization records), and brief summaries of the relationships among important terms (subject records). This contextual information would give searchers strategic tools, clues to help them find useful images (Bearman, 1989, Winter; 1989, Summer)..

Matrix

Initially I attempted to attach appropriate fields to each of the five types of records identified above. It then occurred to me that by generalizing some fields and by applying only selected fields from the total group to any given record type, I could design a matrix of fields and record types that would take care of our requirements more efficiently (Oddy, 1991). The project's indexing committee took my initial draft of the matrix through many revisions before settling on the version printed here (see Figure 3).

Committee Members included Roger Hiemstra, Barbara Florini, Rae Rohfeld, Terry Keenan, and myself as chair.

Figure 3, Data Model Matrix.

Field Name	Collection	Set	Org. Vita	Person Vita	Subject
Item number	•	•	•	•	•
Record type	Collection	Set	Organization	Person	Subject
RLIN ID	•	(from coll.)			
Main Entry	Coll. name	(from coll.)	Org. Name	Pers. Name	Key word
Title	•	(from coll.)			
KLARS source			Who created	Who created	Who created
Physical description	Linear ft.	size & type			
Search strategy hints	•		•	•	Broader.narrow
Bio/historical note	•		•	•	
Summary description	Content	Content	Mission, func, struc		Scope note
Series	•	•			
Folder title		•			
Set title		•			
Begin date	Coll. coverage	Documents	Founding	Birth	
Source begin date			•	•	
End date	Coll. coverage	Documents	Cessation	Death	
Source end date			•	•	
Form/genre	•	•			
Education				•	
Work history				•	
Assoc. subjects	•	•	•	•	Related terms
Assoc. people	•	•	•	•	
Assoc. organizations	•	•	•	•	
Assoc. events/projects	•	•	•	•	
Assoc. places	•	•	•	•	
Publications	Runs in coll.	Within set	Produced by	Written by	
References to	Pubs based on		Pubs describing	Biographies, etc.	
Location	Syracuse U.	Box, folder	•		
Provenance	•				
Restrictions	•				
Preferred citation	How to cite KLARS	•	•	•	•
OCR'd text	•	•			

Across the top of the matrix are the five record types: collection, set, person, organization, and subject. Down the left side are the thirty-two possible fields. Only five fields show up in every record type, and record types employ a maximum that varies from eight fields (for subjects) to twenty-five fields (for collections). A few fields are required; many can be used as needed.

Fields and field labels were abstracted in the process of creating the matrix. For instance, the *begin date* field shows the earliest date of papers in a collection, the date of the first item in a set of documents, the birth date of a person, or the founding date of an organization. In cases where this variation is not self-explanatory, brief notations about the differences are included in the matrix.

Each record type shows one "view" of the data, using one subset of the fields in the matrix. For each view we worked out detailed rules for choosing field content, and standards for data entry (Data Record Views and Input Standards, 1991, May).

Of the five kinds of records, those for whole collections were the easiest to define. The archival community spent several years developing a standardized "container" to facilitate exchange of computerized archival collection records. The result of their work, the MARC/AMC format, (MAchine Readable Cataloging /Archives and Manuscripts Control) provided us with a model (Matters, 1991). At Syracuse, a MARC record is prepared for every newly-processed manuscript collection, and such records are then entered into the RLIN (Research Libraries Information Network) database. This allows people to locate collections nationwide through online searches. We modified the format to fit our purposes, but most of the data in our collection records was taken directly from associated RLIN records. Facsimile images of related finding aid pages were linked to the collection records. The print quality of finding aids is generally quite good, so the bitmapped images of pages were translated, using OCR, into machine-readable text, which was then stored in a searchable *contents* field in the record for the associated collection.

Set records held information about small groups of documents, usually folders. They always carried basic identifying information from their associated full collection records, so users could keep track of the collections from which sets of documents came. When print quality was good enough, OCR translations of document facsimiles were made and kept in content fields. When OCR was not

feasible, the structured record was the only link to document content. All fields except "content" described the group, not individual documents. In no case would we use this structure exhaustively, but when it seemed important to note, for instance, that a particular person figured prominently in a group of documents, we had a place to put that information.

The contextual record types are the most unusual part of our data structure. Person and organization records summarized personal or organizational histories and made note of other people, organizations, events, projects, and places that figured in those histories. Sources of the data collected in these records were noted. Subject records were taken from the *Kellogg Adult Education Thesaurus,* (Allen, 1991) and showed relationships among terms in the notoriously confusing vocabulary of the field. These three kinds of records could potentially become an extended variation on authority files, records kept by libraries that show preferred forms of names and words. Researchers could gather considerable understanding of the adult education domain by exploring them.

An unusual field, *strategy hints*, allowed us to modify records with additional information learned as the documents were used by researchers. Suggestions for approaches to the documents; links between people, organizations, and collections; notes about location of other relevant collections--many kinds of data--could be entered into this field.

USING KLARS

Data Capture

Machine readable data could be entered into KLARS in three ways: direct keying, optical character recognition, and import of files from other machines. Descriptive information about collections and sets of documents was entered directly by library staff members. Content of finding aids (for collections) and, when possible, content of individual documents (for sets) was translated into machine readable form and brought into KLARS using optical character recognition. Data for the three contextual-record types (person, organization, and subject) was created on a variety of computers and imported into KLARS electronically.

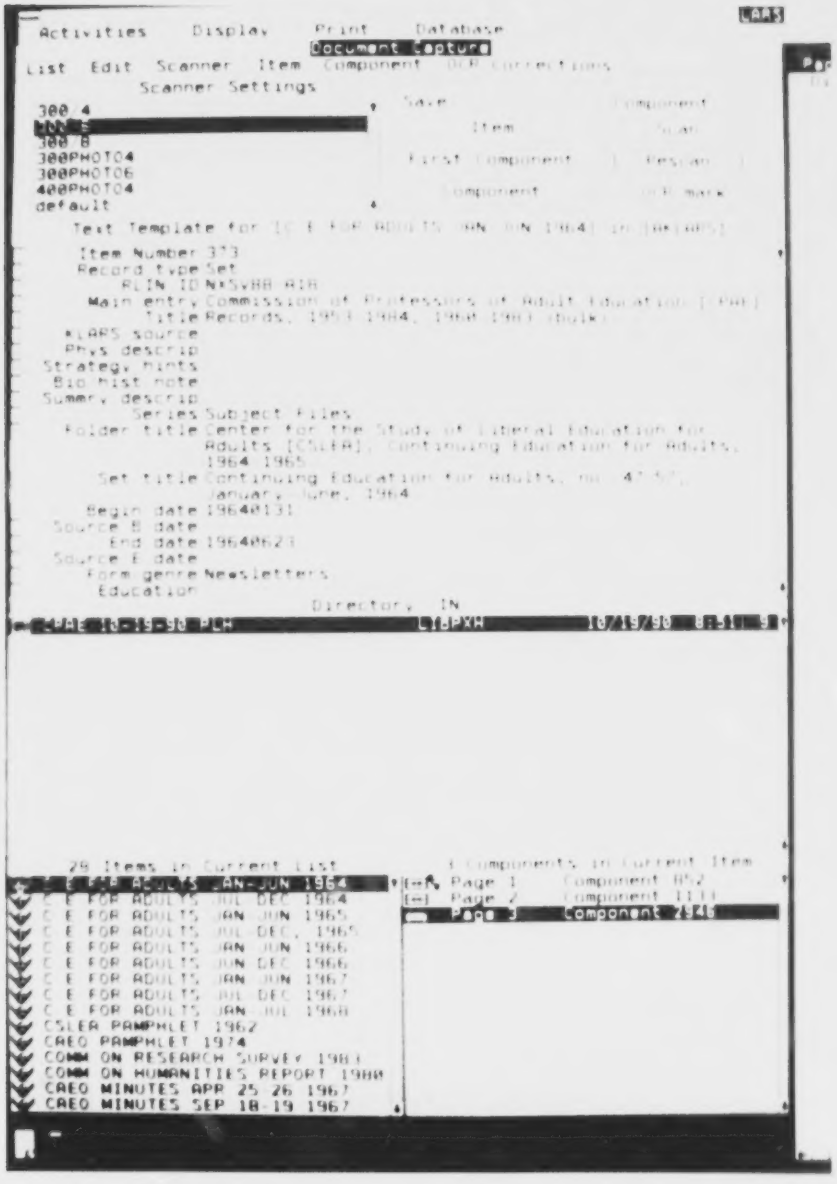
When documents were entered into KLARS, staff members would first look through the papers and complete forms that helped them organize data for input. The process and content differed depending on whether the material to be entered would form a collection record or a set record. After the data was ready, the staff person would type the information he or she had collected into the system's data entry template and save it as what we called an *item*. An item, then, was a machine readable description of a group of documents, or of a person, organization, or subject. Figure 4 shows a document capture screen, with the item text in the text template box. Note that the entire list of thirty-two fields was available to the staff member inputting data. The template box scrolled to show the remaining fields.

Next, digitized images were scanned in for all pages associated with the descriptive information. Each image became a *component* in KLARS. The system provided quite a few scanning options, allowing input staff to optimize the images they captured (see Figure 5). Image bitmaps were compressed to save storage space without losing data, and then stored on optical Write Once Read Many (WORM) disks. WORM technology was chosen because it creates a permanent record (once the data is entered on the disk, it cannot be changed) and because the disks can hold huge amounts of information.

When appropriate, staff members would then send the images to the optical character reader for translation and, if necessary, would edit the results. OCR translations are inevitably imperfect, but editing the files is time consuming. We decided to edit only the finding aids for collections and to let the results of OCR on individual documents stand.

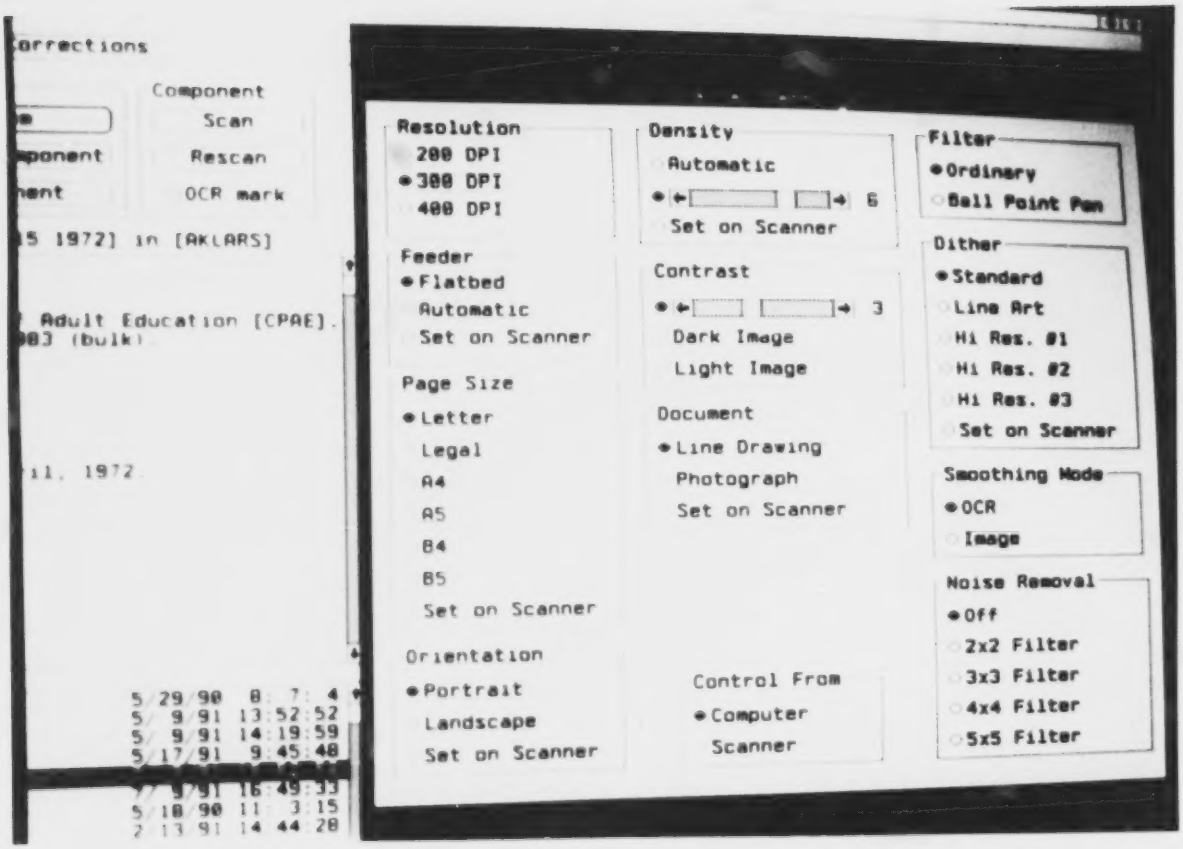
After a large number of items and components were entered into the system, staff members would send the data to the indexer, a program that created an *inverted file,* an alphabetical list of every substantive machine-readable word present in the descriptions. This file noted not only each word, but also where each instance of each word was located, thereby allowing the system to find, very quickly, any instance of any word, i.e., to do full-text searching.

Figure 4. Document capture screen.



BEST COPY AVAILABLE

Figure 5. Scanning options.



Roger Hiemstra developed most of the person records. Another faculty member, Rae Rohfeld, developed the organization ones; and a field worker from the School of Information Studies, Eileen Allen, developed the thesaurus and the subject records. When all of these had been completed in the appropriate format, they were simply imported into the KLARS system. Each record became an item. Images (components) were not associated with person, organization, or subject records.

Search Mode

Researchers could work in search mode only. Searchers spent most of their time doing three things: searching for and looking at machine readable descriptions of

BEST COPY AVAILABLE

documents or contextual information, looking at facsimile bitmapped images of documents, and taking notes in a rudimentary word processor available to them in one window on the screen.

KLARS presented to the user a search screen that allowed (a) a template search or (b) a search using commands. Template search was designed to simplify the process for users who did not fully understand principles of Boolean searching, which employs the logical operators AND, OR, and NOT. These operators can, when well used, give relatively good results in the search for topical information. Used poorly, they make the search very inefficient and misleading. By giving users access to a screen listing fields and letting them fill in the blanks with words they wanted to find, we hoped to make it easier for beginners to use the system. They could restrict the search to appropriate fields if preferred. The system showed them the implicit logic it was using when it performed the search. We could modify this implicit logic as we tried to find the most generally useful combinations of logical operators for our search template. More experienced users could use the command line directly to create appropriate Boolean expressions. Command searches skipped the template completely.

A third approach, Build Command, gave users the chance initially to specify a template search, then to modify the result. In this case, a user would not be responsible for the entire creation of a command line strategy, but could doctor the one his or her template search created.

When a search produced results, a researcher would see a list of items that satisfied the search criteria. Remember, items are the records in KLARS. When they are records for a collection or a set, each item describes a group of documents, perhaps a very large group. By clicking on an item label, the user could bring up a window showing the associated descriptive information, with search words highlighted (see Figure 6). Whenever one item was being displayed, the item list would also show a list of associated components, or images. If the user wanted to see an image linked to a particular description, he could click on that component identifier, and a window showing the image would appear (see Figure 7). Researchers could browse through all the images in a given item, only a few, or could just look at the machine readable item descriptions themselves.

Figure 6. An item (descriptive information), with search words highlighted.

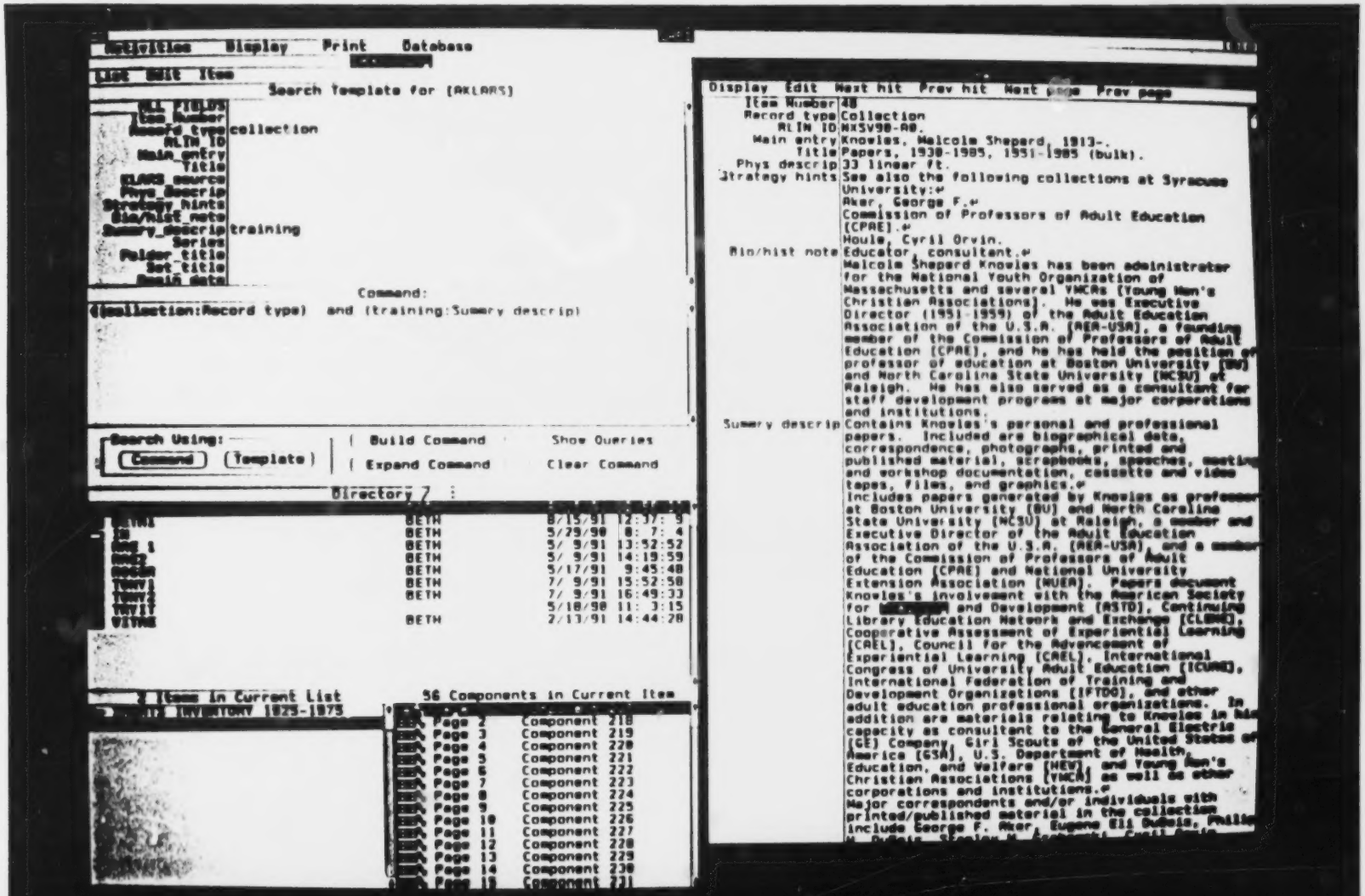
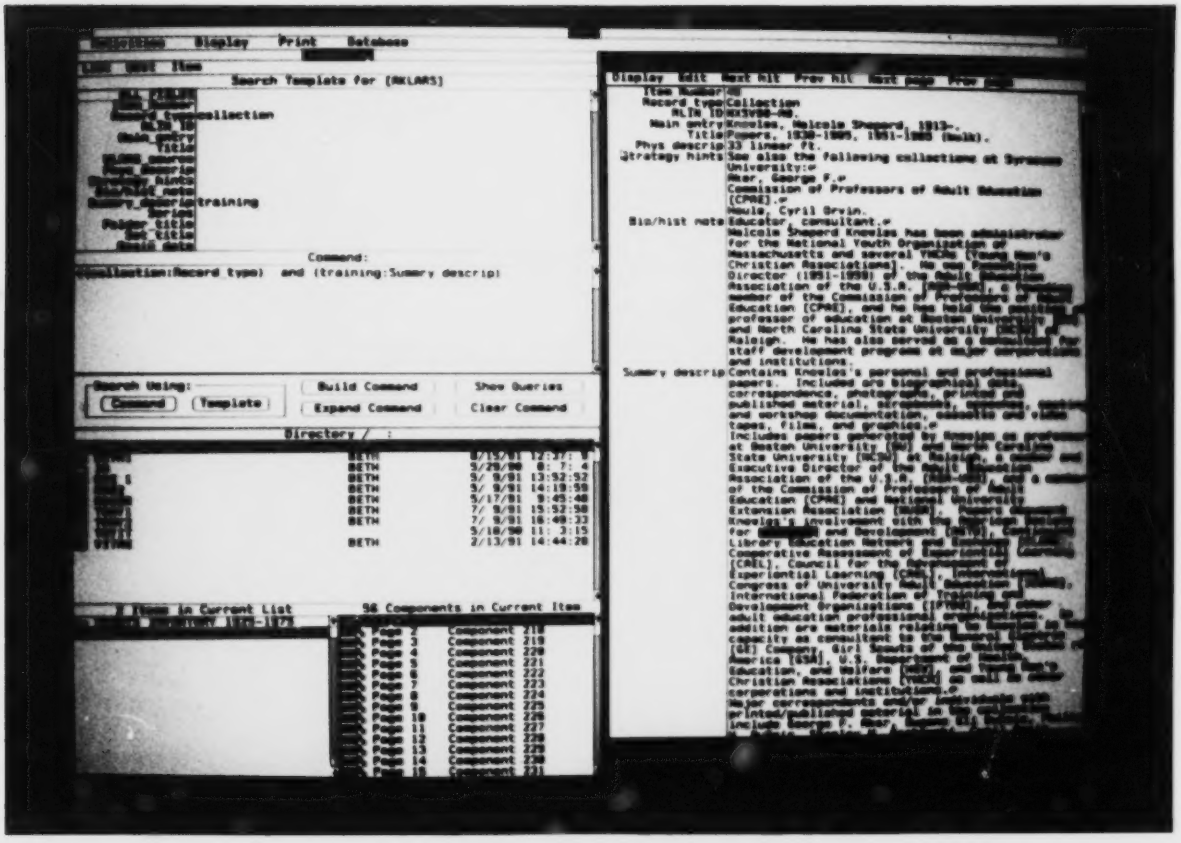
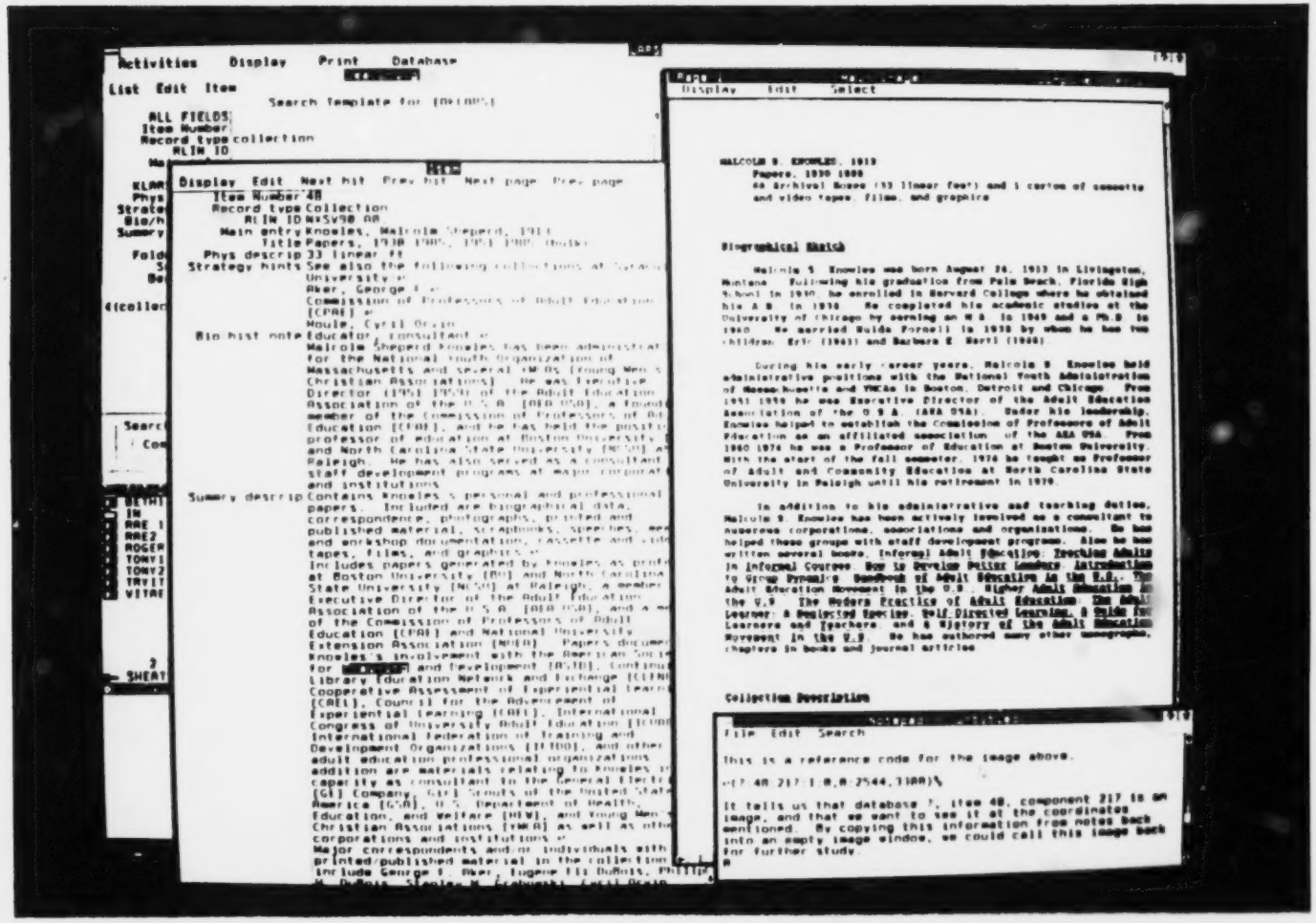

Display Edit Next hit Prev hit Next page Prev page
Item Number 48
Record type Collection
RLIN ID NXSV98-A8.
Main entry Knowles, Malcolm Shepard, 1913-.
Title Papers, 1938-1985, 1951-1985 (bulk).
Phys descrip 33 linear ft.
Strategy hints See also the following collections at Syracuse University:
Aker, George F.
Commission of Professors of Adult Education (CPAE).
Houle, Cyril Orvin.
Bio/hist note Educator, consultant.
Malcolm Shepard Knowles has been administrator for the National Youth Organization of Massachusetts and several YMCAs (Young Men's Christian Associations). He was Executive Director (1951-1959) of the Adult Education Association of the U.S.A. (AEA-USA), a founding member of the Commission of Professors of Adult Education (CPAE), and he has held the position of professor of education at Boston University (BU) and North Carolina State University (NCSU) at Raleigh. He has also served as a consultant for staff development programs at major corporations and institutions.
Summry descrip Contains Knowles's personal and professional papers. Included are biographical data, correspondence, photographs, printed and published material, scrapbooks, speeches, meeting and workshop documentation, cassette and video tapes, films, and graphics.
Includes papers generated by Knowles as professor at Boston University (BU) and North Carolina State University (NCSU) at Raleigh, a member and Executive Director of the Adult Education Association of the U.S.A. (AEA-USA), and a member of the Commission of Professors of Adult Education (CPAE) and National University Extension Association (NUEA). Papers document Knowles's involvement with the American Society for [redacted] and Development (ASTD), Continuing Library Education Network and Exchange (CLENE), Cooperative Assessment of Experiential Learning (CAEL), Council for the Advancement of Experiential Learning (CAEL), International Congress of University Adult Education (ICUAE), International Federation of Training and Development Organizations (IFTDO), and other adult education professional organizations. In addition are materials relating to Knowles in his capacity as consultant to the General Electric (GE) Company, Girl Scouts of the United States of America (GSA), U.S. Department of Health, Education, and Welfare (HEW), and Young Men's Christian Associations (YMCA) as well as other corporations and institutions.
Major correspondents and/or individuals with printed/published material in the collection include George F. Aker, Eugene Eli Dubois, Phillip
Search Template for (RLIN)
collection
training
(collection:Record type) and (training:Summry descrip)
Search Using:
Command
Template
Build Command
Expand Command
Show Queries
Clear Command
Directory 7
56 Components in Current Item

Figure 6. An item (descriptive information), with search words highlighted.



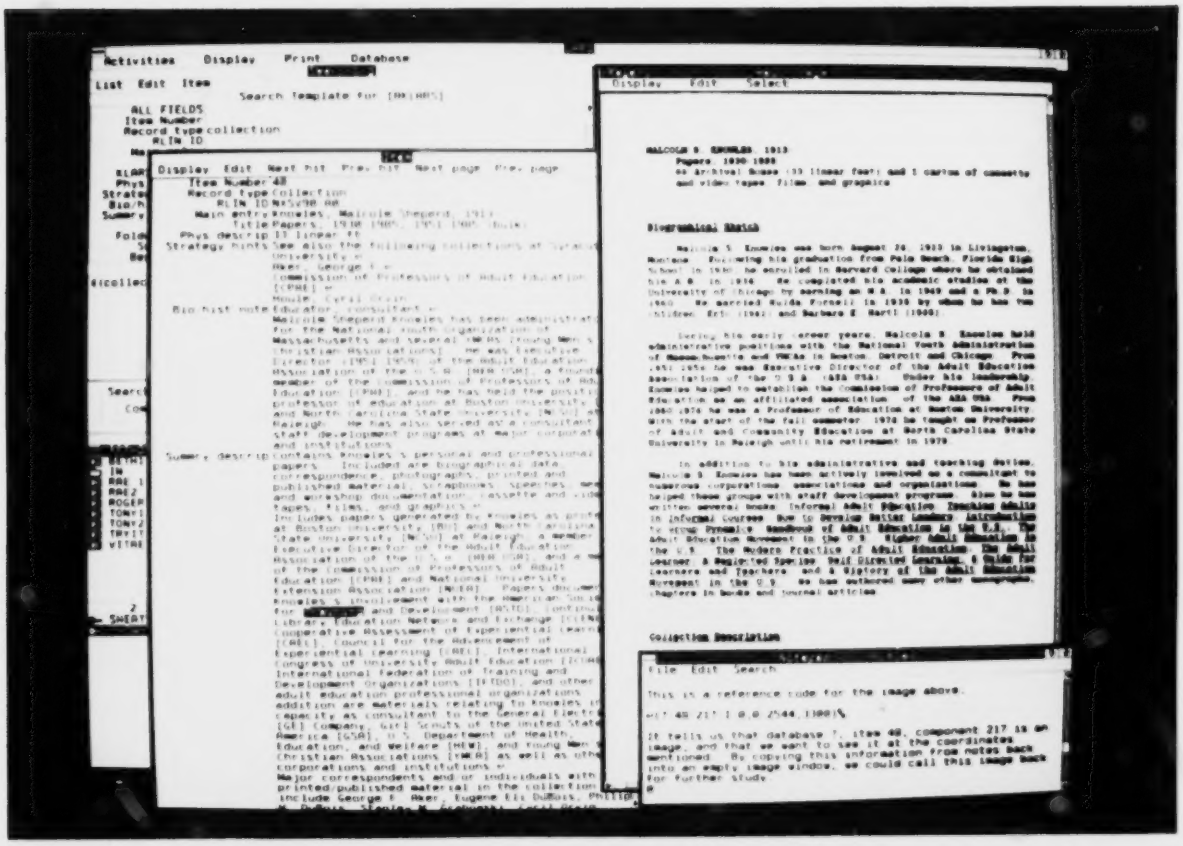
If something particularly interesting struck a researcher, he or she could open up a small word processor, the notepad shown in Figure 8, and take notes. Descriptive information (machine readable) could be cut and pasted into these notes, but images could not. However, a user could copy coded references to the images of interest into his or her notes, take the notes away at the end of the session on a floppy disk, and then recall images from the references at the next KLARS session. Small printers (for machine readable text only) attached to the workstations made it possible for researchers to print out drafts of their notes.

Figure 7. An image (component), linked to a given item (description).



BEST COPY AVAILABLE

Figure 7. An image (component), linked to a given item (description).



When images were of sufficient interest to researchers, they could create lists of desired items and send the lists to be printed. Good, laser-printed facsimile copies could thus be made easily and taken away for study. Associated with each UserID was a small amount of magnetic storage space in the KLARS system. Users could name and store lists of documents or records to which they wanted to return.

Networking

As originally envisioned, KLARS would have been available to anyone who had access to similar equipment (a PC with image handling capability and a high resolution screen) and a connection to a high-speed telecommunications network.

We were unable to implement this aspect of the work fully, although we did link the P-95 to the campus network, which meant that the back end data could be accessed widely. In addition, we set up one workstation across the campus from the rest of the system, images and all. Because of the system design, frequent calls to the back end were a given, and each call from the distant workstation implied trips back and forth through the network. There was no noticeable degradation in performance in the distant workstation.

MOVE TO MACKLARS

The project ended on August 31, 1991, and the university chose not to support the KLARS hardware and software. In order to preserve our work, we developed a HyperCard system, named MacKLARS, on a Macintosh computer. All of the machine readable data captured in KLARS were transferred into this system. Unfortunately we lost the images. However, MacKLARS is capable of holding them, and if additional money can be found someday, we would like to complete the model by replacing its lost images. Faculty, staff, and students in the adult education program, and special collections staff in the library are interested in the replacement system (with or without images).

MacKLARS, like its predecessor, lets researchers search for any machine readable word, whether fielded data or OCR'd text. We used the commercial program HyperKRS as a search engine. It is designed to work with HyperCard software. This combination solved the problem that caused us to attempt custom software development for the original system--blending a database capable of holding images with text retrieval; and the user interface for this Macintosh program is much easier to learn and use.

HyperCard includes, in effect, a very high level programming language (HyperTalk), which makes it possible for programmers without extensive Macintosh experience to program multimedia applications. The most recent revision of HyperCard (version 2.0) and HyperKRS were combined into our customized application by a Kellogg Project staff member, system administrator Yuming Tung. The whole process took approximately six months.

After data transfer to MacKLARS, we had a record for every adult education collection in the George Arents Research Library for Special Collections. Within all but two of these records the contents field holds a machine readable copy of the finding aid text, fully edited. Set records for groups of documents from the Commission of Professors of Adult Education collection have also been transferred. Only some of these have been OCR'd, and none of the translations have been edited. In addition there are records for small numbers of people, organizations, and subjects. Roger Hiemstra intends to keep developing these contextual records, so the numbers should increase.

SUGGESTIONS

As pathfinders in the quest for image systems useful in research, I think we owe our successors a few comments on what we have learned. The following suggestions will point out some pitfalls and danger points:

• Do your best to learn about the vendor from which you purchase a system. Even when this is done, there are big risks with new technology. We carefully sought recommendations from others for whom Plexus had contracted to build similar custom systems, and the results were very positive. The company was considered one of the best in surveys and articles. The bankruptcy, caused by the withdrawal of support by venture capitalists, was as big a surprise to our programmers as it was to us. It was also by far the most devastating blow sustained in the development of KLARS. Extensive delays during completion-contract negotiation were most noticeable at first; over time the diminished size of the programming team, reduced from five full-time people to one part-time person, and the lack of long-term software support were most damaging.

• Consider carefully what staff will be needed to manage a new system. We are all accustomed to off-the-shelf hardware and software, and office systems that require limited in-house expertise to manage. A large system using new and rapidly growing technology and custom software is a completely different matter. Those who wrote the Kellogg proposal had not anticipated the need for a full-time system administrator for KLARS. We found ourselves with insufficient support for the system until rather late in its development, May 1990. Perhaps some of the bugs

could have been fixed more quickly had we had more expertise on the project earlier.

• Think ahead and plan for change. When new to this environment, it is hard for one to imagine how very quickly technology changes. Improved hardware and software become available almost before you have learned how to use what you have. Machines purchased in 1987 are now out of date. Provision for updating equipment and migrating software to new machines was not made in the original Kellogg proposal. To keep a system like KLARS current is an expensive proposition, but a necessary one if you expect to keep the system alive.

• Although one has only limited control in an environment not yet settled into standardization, two tactics stand out as being particularly important if you want to keep options open when purchasing a system: (a) Look for systems that employ open architecture; and (b) allow for flexibility in system configuration, rather than committing to proprietary products. Whenever a standard exists, insist on following it.

• If possible, try to avoid contracting out for custom software development. It is much more difficult to keep communication lines open and to maintain control over programming activities when they are not part of your organization. This became one of our biggest problems. When feasible, using off-the-shelf software is clearly best, but one cannot always find what one needs. If custom work is indicated, it would certainly be preferable to have programmers on your own staff, where they would be easily accessible and responsible to the same bosses. When Plexus Computers collapsed, our electronic mail link died. Communication with HCI and the sole remaining programmer was difficult, and the distance between Syracuse and Santa Clara became a gulf.

• Consider carefully what method to use for storage when permanence is desired. Optical disks open up some remarkable possibilities. On the other hand, the disks themselves are of uncertain longevity. Perhaps more importantly, they require technical assistance from evolving machines before they relinquish the information stored on them. A replication schedule for copying digital information onto new disks at regular intervals is a must in this environment. Even if the disks maintain

their integrity for 100 years, it is unlikely that equipment 100 years from now will be able to read them.

• Be wary when committing yourself to a technology that is not yet standardized. There are no standards yet for formatting data on WORM disks. When we began to plan the MacKLARS system, we found ourselves unable to salvage the images we had scanned into KLARS. They had been stored in a format tied to the Plexus hardware. Images on Macintosh machines are stored in a completely different format. This lack of standardization is a serious problem, and discourages people from entering the image computing arena. Standards have been agreed upon internationally for CD-ROM disks, so they can be read by many kinds of machines. Perhaps WORM disks will reach this point in the near term.

• Plan for adequate funds. It is easy to see the need for money to purchase initial hardware. It is harder to remember that upgrades may be necessary before very long. If software can be purchased completely off-the-shelf, this may be a relatively predictable and moderate expense. Unfortunately, optical systems incorporating full-text retrieval are rather unusual and may require considerable adaptation beyond available commercial programs. As new products come to the market, this should become a smaller problem. Nevertheless, software costs in terms of money, staff, and time can be substantial. When the proposal to the Kellogg Foundation was written in 1986, the Syracuse team did not fully comprehend the immensity of its custom software needs, and therefore did not budget adequately for them.

Expenses other than for hardware and software are perhaps less obvious. Maintenance of hardware can be expected to cost approximately ten percent of the original cost each year. For an expensive system, this can be an arresting amount of money. Design and implementation of a new system usually implies the addition of new staff members, or at least a reshuffling of duties. Development of a training program for users of the new system is another source of hidden costs. Even the simple disruption of routine implied in rethinking processes and learning new ways of doing things has serious cost implications.

• Do not get mentally locked into a failing effort. It was only when we really gave up on the original KLARS system, after fighting its loss fiercely, that we were able to

see our way to the MacKLARS phoenix. By anyone's assessment, the new system is faster, more stable, and far easier to use.

• Consider legal implications. Although we were lucky enough not to become a test case, anyone working in this area needs to keep a weather eye on developments in copyright law. The system we dreamed of could have become a case for debate between those supporting ownership of intellectual property rights and those supporting improved access to research materials. It is difficult to find a mutually satisfactory path between these competing interests, and copyright law is only beginning to deal with the new wrinkles brought by the advent of image systems. In the next few years there will no doubt be test cases in which some of the rules of the game are set down. Until then, caution is in order. Similarly, privacy law may affect some systems. Is too much information about individuals or organizations being made available too widely? Third, consider carefully how to ensure the security of data. Who will have access to the system? What controls will be kept over use?

• Consider whether you want your data to be exchangeable among libraries and information centers. The link between our data matrix and the MARC/AMC format is probably not close enough for a system that exists in the current library/information world. Certainly, collection records, and probably set records as well, could be framed in MARC style. David Bearman and others have urged the extension of authority files into reference files that could be made available to scholars as well as processors (Bearman and Szary, 1987). As one of the designers of the AMC format, Bearman tried to make the container broad enough to hold such files. Anyone doing additional work along the lines we have begun might want to consider whether it is possible to develop necessary structures for contextual fields within the MARC framework.

• Last but not least, allow enough time! Double and triple initial optimistic estimates. Especially during the period in which systems like KLARS are young, no one really knows all the answers. Everyone involved in this kind of work faces delays and finds unexpected problems.

CONTRIBUTIONS

As image systems become more widespread, mechanisms for giving access to the images will be refined. Most commercial systems use rather rigid database management structures to store machine readable data related to images. Systems intended for use by researchers do not fit easily into such structures. Research questions are far less predictable than those that occur in a records management environment, for instance, and require much more flexible approaches. The mixture of text retrieval and image management that we implemented is a harbinger of things to come. We presented a design meant to facilitate research, and one that would remove the tight physical link between repository and data.

Commercial systems that combine these elements are just beginning to arrive on the scene. Even when they become easily available, people will still need to decide how to organize their data. It is here that our work may be able to show the way. Our solution to the problem of giving access to specific images while making use of traditional archival processes may provide implementation suggestions to others. Important elements of our solution include:

1. Using a form of applied indexing on groups of documents, and combining that with automatic indexing of individual document content to give both breadth and specificity to online document access.

2. Creating records that describe the context in which documents were created (people, organizations, subjects), and keeping those separate from but available with records that describe the documents themselves.

3. Using unedited OCR'd versions of text from images as rough indexes to those images.

CONCLUSION

It takes time for the possibilities inherent in new technologies to become evident. Our initial human tendency is to do the old jobs in the new environment for a while; only later do people begin to see options that make formerly unimaginable things possible. I think that image systems will eventually move in fresh directions, and new techniques will be needed. Perhaps what we have done will suggest welcome ideas to those who follow.

REFERENCES

Allen, E. (1991). *Kellogg adult education thesaurus* (Technical Report No. 5, 2nd ed.). Syracuse, NY: Syracuse University Kellogg Project. (Available from ERIC Document Reproduction Service as of October 1991)

Bearman, D. A. (1989, Winter). Archives and manuscript control with bibliographic utilities: Challenges and opportunities. *American Archivist, 52*, 26-39.

Bearman, D. A. (1989, Summer). Authority control issues and prospects. *American Archivist, 52,* 286-299.

Bearman, D. A., & Szary, R. (1987). Beyond authorized headings: Authorities as reference files in a multi-disciplinary setting. In K. Muller (Ed.), *Authority Control Symposium* (ARLIS Occasional Papers No. 6, pp. 69-78). Tucson: Art Libraries Society of North America.

Bearman, D. A., & Lytle, R. H. (1985-86, Winter). The power of the principle of provenance. *Archivaria, 21,* 14-27.

Data record views and input standards. (1991, May). In J. Hugo (Ed.), *Kellogg Library and Archive Retrieval System Document Capture Manual* (Appendix A, pp. 104-121). Syracuse, NY: Syracuse University Kellogg Project. (Available from ERIC Document Reproduction Service as of December 1991.)

Dowler, L. (1988, Winter/Spring). The role of use in defining archival practice and principles: A research agenda for the availability and use of records. *American Archivist , 51*, 74-86.

Keenan, T., & Oddy, E. C. (1988, November). Developing an optical disk system for adult education manuscripts: The Kellogg Project at Syracuse University. In D. L. Blamberg, C. L. Dowling, and C. V. Weston (Eds.), *Proceedings of the United States National Agricultural Library's Conference on Applications of Scanning Methodologies in Libraries* (pp. 79-96), Beltsville, Maryland: National Agricultural Library.

Lytle, R. H. (1980a). Intellectual access to archives: I. Provenance and content indexing methods of subject retrieval. *American Archivist* , *43* , 64-75.

Lytle, R. H. (1980b). Intellectual access to archives: II. Report of an experiment comparing provenance to content indexing methods of subject retrieval. *American Archivist* , *43,* 191-207.

Matters, M. (1990). *Introduction to the USMARC format for archival and manuscripts control.* Chicago: Society of American Archivists.

Oddy, E. C. (1991, October). Data structure for an archival image system: A proposal. Paper to be presented at the American Society for Information Science Special Interest Group on Classification Research Workshop, Washington, DC.

Pugh, M. J. (1982, Winter). The illusion of omniscience: Subject access and the reference archivist. *American Archivist* , *45*, 33-44.

Rohfeld, R., Keenan, T., & Oddy, E. C. (1990). Introduction to archival research: How are libraries and archives different? Syracuse, NY: Syracuse University Kellogg Project. (ERIC Document Reproduction Service No. IR 053 272.)